

Facilitating the Transition from Manual to Automated Nurse Rostering

Received: date / Accepted: date

Abstract After several decades of academic research in the field of automated nurse rostering, few results find their way to practice. Often, the configuration of a software system for automated rostering presents a task considered too time-consuming and difficult. The present article introduces a methodology for automating part of the costly and unintuitive configuration process by automatically determining the relative importance of soft constraints based on historical data. Naturally, this automated approach can only be reliable in the absence of transient effects and drastic changes. The approach is evaluated on retrospective and prospective case studies, and is validated by health care practitioners partaking in an experimental study. The results show that, given relevant historical data, the presented approach simplifies the transition from manual to automated rostering, thus bringing academic research on nurse rostering closer to its practical application.

Keywords nurse rostering · weighted sum objective function · automated weight extraction

Introduction

Over the past forty years, academic research on nurse rostering has facilitated health care practitioners in allocating nurses to shifts (Hulshof et al, 2012). Manually designing rosters is complex, labor-intensive, and consequently may yield poor work schedules for personnel. Automated scheduling reduces the time and costs incurred by manual planners and has the potential to produce high quality sched-

ules improving contentment and effectiveness of the workforce, and thus personnel retention (Kellogg and Walczak, 2007; Tourangeau et al, 2006).

Despite the advancements of automated nurse rostering techniques, scheduling by hand is still standard practice in many hospitals. Kellogg and Walczak (2007) have explored the extent to which academic research has actually been applied, and identify a large research-application gap. An important cause of this gap concerns the transition from manual to automated rostering. Many hospitals consider the implementation of an automated rostering approach too costly or too difficult. While the nurse rostering literature focuses mainly on automatically constructing rosters, it rarely addresses the implications and difficulties of actually implementing an automated rostering technique in a hospital environment. Expert knowledge is needed for defining all departmental rules, hospital regulations and staff preferences, as well as their relative importance (*weights*), in an automated rostering system. Configuring the software is time-consuming, as it involves translating skills and expertise in numerical form, which is not intuitive, even for experienced planners. Many health care practitioners see this time investment as a deterrent to the use of automated nurse rostering software and prefer to continue rostering manually (Kellogg and Walczak, 2007).

These issues are confirmed by Petrovic and Vanden Berghe (2012), who argue that correctly modeling the problem is critical in practical nurse rostering. They evaluate two approaches to automated nurse rostering, based on seven criteria. The *expressive power* criterion refers to ‘the ability of the model/system to represent a wide variety of real world constraints and rostering characteristics.’ The concepts of constraints and weights strongly support this modeling requirement. However, as Petrovic and Vanden Berghe (2012) mention, the need to manually prioritize constraints is a severe drawback.

In modern hospitals, much of the domain knowledge needed to configure a software system for rostering is already available in the form of historical data. The present article explores the possibility to use these data for partially automating the configuration of a rostering system, in order to facilitate the transition from manual to automated nurse rostering. A methodology is proposed for extracting constraint importance, in the form of weights, from past rosters. The approach has been named Automated Weight Extraction or AWE. This method has been empirically evaluated on real world data and with the help of health care practitioners in two case studies of Belgian hospitals. The results show that, given an appropriate set of historical data, the proposed AWE method is capable of automatically extracting weights suitable for automatically producing rosters that are usable in practice.

The remainder of this article is structured as follows. First, the difference between the focus of the current academic approach to rostering and the actual practical approach is outlined. The rostering task is classified according to the level of automation in its two main phases. Then, the AWE approach is presented in detail. Data from two real cases are used to analyze and empirically evaluate the AWE approach. The last sections conclude the article and identify a number of directions for future work.

Nurse rostering: theory vs. practice

Nurse rostering is a personnel scheduling problem in health care constituting a particularly challenging task compared to other personnel scheduling domains (Ernst et al, 2004). Hospitals operate around the clock, introducing specific constraints on rest times and on night and weekend shifts. The problem is further complicated by a large range of shift types and skill requirements. The planning horizon typically spans a few weeks to a few months, with staff requirements varying each day.

The present contribution focuses on departmental rostering, where the head nurse of the ward is responsible for constructing the rosters. Rostering is performed accordingly in the hospitals considered in the case studies. Team rostering and self rostering are two other rostering policies, where decision making is delegated to groups or to individual nurses, respectively. These techniques increase the perception of autonomy and reduce the rostering effort of the head nurse. They are restricted to small and medium-sized wards (Silvestro and Silvestro, 2000).

The academic operations research literature on nurse rostering focuses mainly on optimizing rosters of health care personnel (Burke et al, 2004). However, in practice, practitioners still need to invest time and effort transitioning to these automated solutions from traditional manual planning. In the following sections, nurse rostering is surveyed, as seen in the academic literature and by practitioners in health care.

Rostering in the academic literature

Nurse rostering is typically formulated as a constrained optimization problem, attempting to satisfy a number of soft constraints (Smet et al, 2013). Feasible solutions need to respect all hard constraints induced by resource restrictions, departmental requirements and contractual obligations. The quality of a feasible solution (the *cost*) is measured by an objective function and is related to the number of respected soft constraints concerning staff and contractual preferences. The large number of soft constraints, often mutually conflicting, makes it

practically impossible to satisfy all preferences. The objective of nurse rostering problems is to minimize the constraint violations. Such optimization problems are thus multi-objective in nature. Ideally, the planner would have to select from a set of Pareto-optimal solutions. Here, a Pareto-optimal solution is one where no constraint violation can be reduced, without increasing another one.

Multi-objective optimization is often addressed by employing a weighted sum objective function, due to its simplicity and ease of implementation. This method scalarizes the vector of objectives into a single objective and enables comparing the overall cost of solutions by obtaining the sum of all violated soft constraints, weighted by their relative importance.

Many nurse rostering approaches in the literature utilize weighted sum objective functions and define their constraint weights with the help of health care practitioners (Abdennadher and Schlenker, 1999; Aickelin and Dowsland, 2004; Burke et al, 2008). Others simply set the weights by trial-and-error, without elaborating on the choice of values or on their effect on the overall quality of generated schedules (Aickelin and Dowsland, 2004; Ásgeirsson, 2014). A combination of these two settings also exists, where health care practitioners define only a general preference ordering of constraints and then researchers choose the numerical values according to some rule (Berrada et al, 1996; Burke et al, 2012). These manual methods of determining constraint weights are time-consuming. Practitioners need to be consulted, or a large sample of tests needs to be performed. In addition, there are no explicit guidelines for setting constraint weights, whereas setting them poorly induces biased solutions. Research has indicated that setting static weights may result in a solution landscape that is difficult to explore (Parr and Thompson, 2007). Different methods have been proposed for modifying constraint weights, such as evolutionary algorithms with stepwise adaption of weights (Eiben and Van Hemert, 1999), local search algorithms which randomly perturb weights (Charon and Hudry, 2001), and other adaptive weight methods (Kelemci and Uyar, 2007). However, these methods rely on additional parameters that still need to be tuned manually.

The weighted sum objective function has some drawbacks. Constraint weights are subjective, need to be explicitly defined and the final solution is sensitive to the chosen values. Numerous other approaches to multi-objective optimization exist, such as the ϵ -Constraint method (Chankong and Haimes, 2008), compromise programming (Ballester, 2007) and evolutionary algorithms (Coello et al, 2007). Most of these approaches maintain multiple Pareto-optimal solutions and select a single solution based on some high level problem information, input by the user. Nevertheless, the weighted sum objective function remains the most preferred multi-objective optimization method. It is used in the majority of the

academic nurse rostering literature (Burke et al, 2004), as well as in practical applications (Mirrazavi and Beringer, 2007; Eveborn and Rönnqvist, 2004; Abdennadher and Schlenker, 1999).

Rostering by health care practitioners

In practice, the rostering task encompasses two phases: the initial investment and the rostering process. Academic research mainly focuses on the rostering process (Ernst et al, 2004), while in practice, the first phase is equally important. This first phase is called initial investment as the hospital needs to invest resources into translating departmental rules and regulations to software constraints and defining their relative importance. Academic research has overlooked this phase, under the assumption that it offers little potential for automated improvement. However, it constitutes an important, time-consuming phase in the transition from manual to automated nurse rostering that needs to be appropriately addressed (Kellogg and Walczak, 2007). Note that this definition restricts the scope of the initial investment to the software configuration. However, in practice, several additional steps need to be undertaken before starting the configuration, such as securing funds, comparing software systems, and negotiating contracts with the software vendor.

Automation of rostering practices

Figure 1 shows an overview of four different rostering practices, classified based on the level of automation of the initial investment and rostering process. The two most automated practices preview future developments. Tasks in non-shaded boxes represent work performed by hand, while the grey boxes indicate automated tasks. The following paragraphs discuss these different practices and highlight advantages and drawbacks.

Fully manual rostering

Many hospitals conduct labor-intensive manual rostering as the transition to automated rostering is considered costly and unintuitive. Even though training personnel for manual rostering is equally costly and may sometimes be more expensive, it follows long-established and implicit procedures, which practitioners are reluctant to replace. Manual planners need to study hospital regulations and departmental constraints before they are able to construct rosters. They need to also learn,

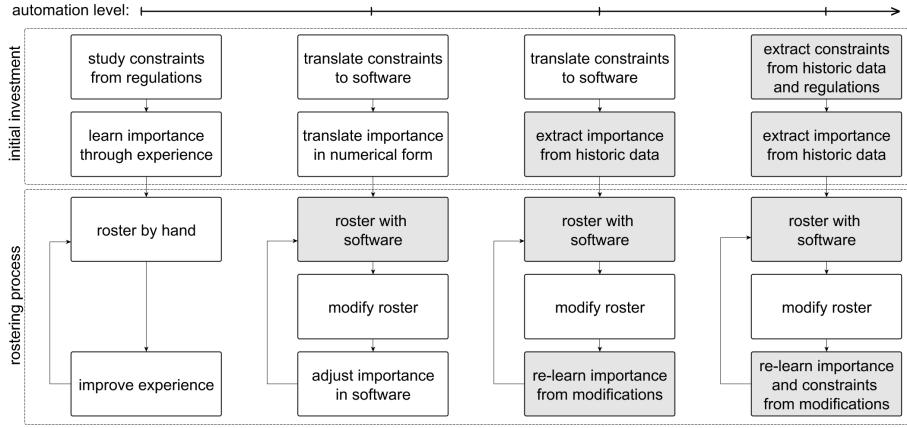


Fig. 1 Overview of different levels of automation of rostering practices and their two main phases. Non-shaded boxes show manual steps, grey boxes indicate automated ones.

through experience, the relative importance of these rules, in order to minimize violations of important constraints.

Departmental rostering puts the responsibility for designing the rosters in the hands of the head nurse, who often has the unrewarding task to trade-off departmental regulations with staff preferences. Nevertheless, manually constructing rosters allows head nurses to tailor rosters to the particular needs of the department based on their experience and familiarity with the personal qualities of the staff. For example, a head nurse may avoid assigning two nurses who do not get along well to the same time period, assign shifts based on personal requests, or assign fewer nurses than required due to a temporary drop in patient admissions. Although handling such implicit constraints may improve staff contentment, it often comes at the expense of the overall organization, which has little control over the final rosters.

Automated rostering with manual transition

Some hospitals have invested in automated rostering software, partially automating the labor-intensive manual rostering while tightening central control (Abdenadher and Schlenker, 1999; Eveborn and Rönqvist, 2004). The initial investment enables automated rostering by allowing different automatically generated solutions to be evaluated and compared. However, as the weighted sum objective function is an approximation of the real implicit roster quality, staff members often manually modify the final solution to correct any perceived faults or to incorporate additional implicit constraints and personal requests.

With no explicit guidelines for translating the importance of constraints, practitioners need to rely on their own experience and intuition when configuring the numerical weights. The manual configuration of the rostering software becomes a costly trial-and-error process, where human experts attempt to correct the search by adjusting constraint importance after observation of the obtained solutions. This iterative adjustment is bound to result in a vicious cycle, as neither the health care practitioners understand the search process, nor can the software interpret the expert's rationale to deliver satisfactory results.

Automated rostering with semi-automated transition

While the previous rostering practice reduces the involvement of experienced planners in constructing the rosters, this article introduces a methodology for further reducing the need for manual input by partially automating the initial investment phase.

Many hospitals aim at increasing the rostering process's automation level by relying on computerized solutions. Such solutions are only as good as the domain expertise they contain. In order to incorporate domain expertise into computerized solutions, either humans need to translate their expertise into a language interpretable by a computer, or the machine needs to extract the expertise of humans. Automated rostering with manual transition addresses the human-computer knowledge gap from the human side first by letting practitioners explicitly translate their knowledge into the language of the automated rostering software.

The present article proposes a new rostering practice where the knowledge gap is tackled from both sides simultaneously. Human experts only define the variables in the objective function, i.e. the constraints, while the machine automatically extracts their relative importance (the weights) from past data. The transition from manual rostering to automated rostering thus becomes semi-automated. The efforts and costs of the transition are partially performed by the machine. This technique reduces the initial investment by supporting the user in configuring the automated rostering software.

In addition to partially automating the initial investment phase, the proposed methodology excludes the need for manual re-adjustment of constraint weights in the rostering phase. The constraint weights are updated in a way similar to the initial phase, namely by examining recent rosters and adjusting the importance of constraints based on the changes introduced by the human planner. Automating the re-adjustment step lowers human bias in the final solution, whereas it has a potential difficulty handling transient effects, such as illness of a staff member or a sudden peak in patient admissions. In addition, drastic long-term changes, such

as re-structuring or merging of wards, may negatively influence the stored domain knowledge and may require re-learning.

Automated rostering with fully automated transition

In the future, a natural extension to the semi-automated transition would be to entirely automate the initial investment phase by extracting not only constraint importance, but also the constraints themselves. The human-computer knowledge gap would then be addressed entirely from the computer side, as the initial investment phase is performed exclusively by the machine. This rostering methodology would allow any manually rostering ward with relevant historical data to effortlessly transition to automated rostering.

Extraction of constraints can be achieved in two distinct manners that can later be combined for improved accuracy. Natural language processing techniques and text mining can be applied to extract information regarding constraints from written hospital regulations and departmental rules. Another method to extract information is to perform pattern detection on historical data using a general knowledge base. This database of constraints can be collected and updated by wards and even shared by hospitals, so that automated methods can detect whether a subset of these constraints has been respected in past (manually constructed) rosters. Once constraints are extracted, their importance can be determined by the method proposed in the next section. As with the above rostering practice, though, automated knowledge extraction may incorporate transient effects and learn constraints induced by temporary roster modifications.

Automated constraint weight extraction

The remainder of this article focuses on automatically defining the importance of constraints in the initial investment phase. It is assumed here that sufficient past rosters are available and that their quality is satisfactory to the department and its personnel. In addition, all necessary constraints have been translated to the automated rostering software. Although the latter is a manual task, the process of adding constraints is rather straightforward, as hospital regulations and departmental constraints are explicitly stated in legal documents. Constraint importance, on the other hand, is implicit and requires experience to learn and translate into numerical weights. The past schedules and constraints need to be available in a structured way, for example in an XML format (Smet et al, 2014).

The proposed AWE approach works as follows. For each constraint, the number of violations is counted in all rosters, as well as the number of times the constraint is

respected. The sum of these two numbers indicates the instances of the constraint, i.e. the number of times the constraint can be identified in all rosters. The new (or automated) weights are defined to be proportional to the ratio of the number of times the corresponding constraint is respected in the past schedules versus the total number of instances of that constraint. Formally, the weight w_c^{new} of each soft constraint c is defined as follows:

$$w_c^{new} = \frac{(n_c - v_c)}{n_c} \quad (1)$$

where n_c stands for the total number of instances of constraint c in all rosters, and v_c is the number of violations of this constraint. The intuition behind this choice is that constraints with a high percentage of violations are less important and should therefore be given low weights. Even though violated constraints might actually be important, they may be difficult (or impossible) to satisfy, e.g. due to conflicts with other constraints. Associating very high weights with such constraints may distort the search process and consume computational resources in vain. Analogously, constraints that are mostly respected in the historical data are either important, or easy to satisfy, and are given a high weight.

It should be stressed that the proposed approach simply extracts information from past data. For this reason, an extracted weight configuration cannot steer an algorithm to generate more preferable schedules than the manual ones, but rather generate schedules that conform to the historical data. The main advantage of the proposed method is that the obtained weights enable automated configuration of the scheduling software and do not rely on the skills of the practitioner to quantify his/her expertise. Moreover, it allows any department with sufficient historical data to switch to automated rostering by allocating only limited resources during the initial investment phase, as the most time-consuming task is automated. Numerous data mining techniques may be applied to extract importance values from manual schedules, such as neural networks (Kudyba and Perry, 2015) and regression (Tai and Machin, 2013). However, the proposed method is intuitive, transparent, computationally inexpensive and above all, simple to explain to health care practitioners without a mathematical background.

A retrospective case study

The weight extraction method was developed when analyzing historical data from a Belgian hospital that has already transitioned from fully manual to automated rostering. The initial investment phase in all wards was performed fully manually. Hospital regulations and departmental rules were translated to constraints in the

automated rostering software, and their importance was set based on the intuition of the head nurse of the respective ward. In this section, three rostering practices are analyzed: fully manual rostering, automated rostering with manual transition, and automated rostering with semi-automated transition. Specifically, the weights and constraint violations in different solutions are compared.

Case description

This case study concerns the reception ward of a modern Belgian hospital with 350 beds. The ward is medium-sized with up to 21 staff members, each working one of three contract types: full time, $3/4$, or another part time. The employees are further categorized according to whether they have telephoning skills or not. Coverage requirements are specified for different types of early, day and late shifts. Besides the contractually required number of hours worked each month, additional hospital-wide and ward-specific constraints are specified in the employees' contracts. The former are concerned with providing sufficient rest time — 35 hours rest per week for each employee, and adequate time between two consecutive working days. The ward-specific constraints restrict certain shift types to be assigned to certain employees (e.g. 4h or 6h shifts should not be assigned to full time employees), and ensure a minimum number of particular shifts per employee. Furthermore, the ward-specific constraints state that an employee should never work two consecutive weekends. Finally, a shift may induce a subsequent number of days off. In total, each roster needs to consider 100 soft constraints of varying importance.

Study design

The hospital provided data in the form of three years of past one-month rosters. These rosters had been constructed manually by the head nurse and served as an initial assignment for the ward before corrections were applied due to illness, unavailability of personnel or reduced workload. These corrections are not included in the analysis, as they are incidental and not representative of the ward's operation.

Three sets of one-month rosters are used in this case study. The first set consists of the rosters manually constructed by the head nurse without support of the automated rostering software. This data set is referred to as the “*Historical data*”. For the same months, two additional sets of rosters are generated using the rostering software system. This system was manually configured by the head nurse with all constraints relevant to the department, as well as the constraint weights. The first of the generated sets, called “*Generated schedules with manual*

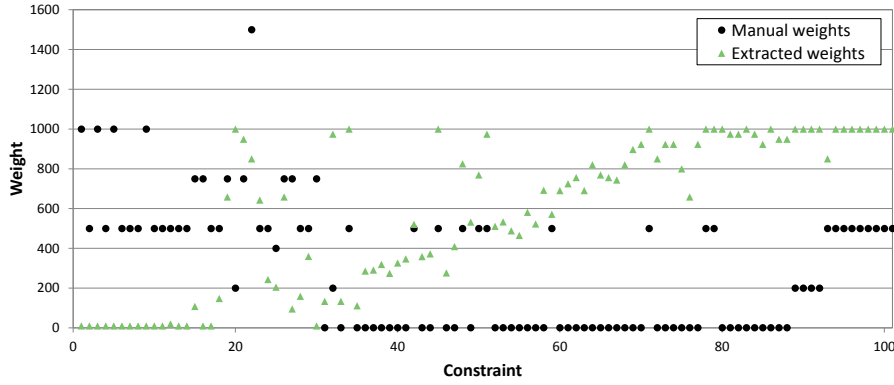


Fig. 2 Manual and extracted weights for the retrospective case study.

weights”, applies the weights defined by the head nurse. The process of automatically generating rosters was repeated for the last set, but instead of the manual weights, AWE extracted weights from the historical data. This was done by, for each constraint, calculating n_c and v_c in each monthly roster, summing all these values, and computing w_c^{new} according to Equation (1). The extracted weights were then used to generate schedules for each of the monthly periods. This last data set is called “*Generated schedules with extracted weights*”.

For analyzing the differences between the three sets of rosters, the following data was collected. Firstly, for each constraint, the weights manually set by the head nurse in the software system were compared with the weights extracted by the AWE approach. Secondly, for each constraint, the total number of violations in all rosters was computed, for all three roster sets. These data was used by the authors to evaluate how the three different types of rosters compare to each other.

The algorithm used to automatically generate rosters is the heuristic search algorithm presented by Smet et al (2014). The allowed calculation time of the algorithm was limited to ten minutes, so as to simulate potential conditions in practice. Mathematical details on the computation of constraint violations are explained by Smet et al (2013).

Results

For all 100 constraints, Figure 2 compares the values of the manually defined weights with the values of the weights extracted by AWE. Although the values range from 0 to 1500, the manual weights were chosen among only a few discrete values, while the AWE approach results in a much more spread out weight distribution. Using a limited number of discrete values for constraint weights may

result in multiple plateaus in the search space and therefore numerous solutions with equivalent search score. A wider range of weight values, on the other hand, results in a less monotone search space that may be easier to explore (Marmion et al, 2011).

The graph also shows that the extracted weights deviate significantly from the manual weights. This suggests that the weights chosen by the head nurse do not correspond with the actual priority of constraints according to the accepted rosters. It is common for head nurses to be satisfied after making changes to the generated rosters, even though the overall cost after such modifications increases (Drake, 2014b). This suggests that the manually defined constraint weights differ from the perceived importance of constraints.

In addition to studying the raw values of manual and extracted weights, the constraint violations in the respective rosters are compared. Figure 3 shows the total number of violations for each constraint in the three sets of rosters. Highlights show the number of violations of “*Important constraints*”, i.e. constraints weighted 1000 or more by the head nurse (cf. Figure 2). Given the assumption that the head nurse was satisfied with the manually produced rosters, the line of the historical data can be interpreted as the preference of the head nurse for trading off constraint violations. Clearly, the constraint violations of the rosters generated with manual weights do not follow the same trend as those of the historical data. This confirms the conjecture that the manually defined weights do not correspond to what the head nurse perceived as important while constructing the rosters. Analogously, this result shows that, although the head nurse most certainly has an implicit knowledge of constraint importance, translating experience into numerical form is not trivial and is bound to produce inaccuracies.

The constraint violations of the rosters generated with extracted weights closely approximate the trend in the historical data. This outcome suggests that rosters generated with the weights extracted by AWE resemble the preferences of the human planner much more than the rosters generated with the manual weights do. This illustrates that the proposed AWE approach succeeds better in determining weights corresponding to the real implicit preferences of the head nurses, than the head nurses are able to manually define.

A prospective case study

Since all wards in the hospital of the first case study had already performed the transition from manual to automated rostering, it was of little interest for both practitioners and authors to let personnel from this hospital evaluate the results of AWE. Therefore, a different Belgian hospital, which recently begun this transition,

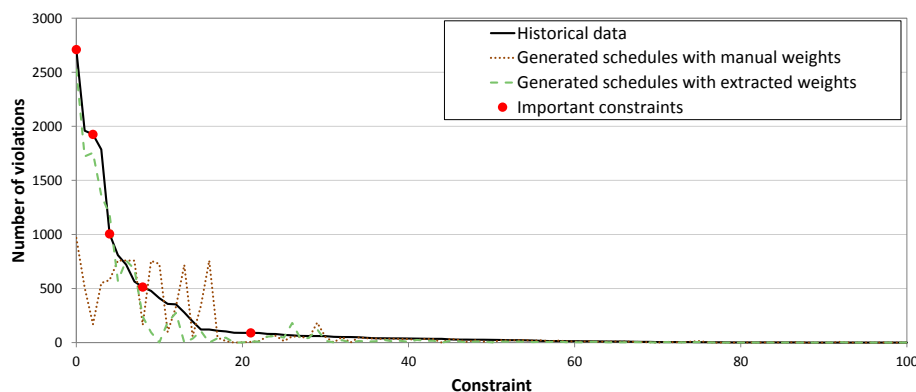


Fig. 3 Comparison of constraint violations in the different rosters.

was invited to validate the approach in practice. In this section, AWE is applied to data from the latter hospital and a similar analysis as in the retrospective case study is performed. Additionally, the results are evaluated with the help of the hospital's personnel.

Case description

This second case study is carried out in a medium-sized Belgian hospital with 400 beds. The hospital has purchased the same automated rostering software as the hospital in the first case study, but only few wards have yet performed the transition. The head nurses of four of these wards have manually incorporated the constraints in the software and have associated a weight to each constraint, based on their experience from manual rostering.

On the one hand, the practitioners were interested in comparing the quality of the newly automated rosters with the ones manually constructed. On the other hand, the hospital was in favor of testing the weight extraction approach in order to facilitate configuration of the software for the remaining wards. Naturally, the authors were also interested in evaluating the AWE approach by comparing the quality of rosters generated with manual weights and automatically extracted weights. Therefore, an experiment was conducted that aimed to evaluate the same three rostering practices discussed in the retrospective case study.

The four wards who agreed to participate in this case study are maternity (MT), general surgery (GS), geriatric medicine (GM) and the operating theater (OT), as displayed in Table 1. Similarly to the retrospective case study, the wards are medium-sized, but now present a mix of care and non-care units. One difference with the first case study is the number of constraints considered when constructing

the rosters. For example, the GS ward specifies no ward-specific constraints. In the MT and GM wards, the number of consecutive working days is limited to maximum five, while for night shifts, the number of consecutive days worked should be between four and seven. The OT ward wants to balance the number of undesirable shifts among nurses in order to avoid uneven distribution of such shifts. All wards however, are subject to the organization-wide constraints, ensuring sufficient rest time between shifts, and forbidding some shifts to be assigned to certain nurses (e.g. 4h or 6h shifts to full time nurses).

The number of constraints in the different wards is furthermore influenced by the number of coverage constraints. The MT ward, for example, has many of these constraints for the different types of shifts.

Table 1 Overview of wards in the prospective case study.

Ward		Employees	Shifts	Skills	Constraints
Maternity	MT	21	7	2	13
General surgery	GS	15	15	2	19
Geriatric medicine	GM	21	17	4	14
Operating theater	OT	23	12	3	19

Table 1 shows a relatively high number of shift types for the GS, GM and OT wards. Typically, there are only four main shift types: early, day, late and night. However, for some of these types, short (4h), medium (6h) and long (8h) variants exist, with different break lengths, which should be assigned to either part time or full time nurses. Table 2 shows the different shift types for the GM ward.

Study design

For each of the four wards, three years of historical data were supplied in the form of past rosters, constructed by the head nurses. The hospital again provided the manually performed software configuration in the form of constraints and their associated weights. AWE was applied to extract new soft constraint weights for each ward, based on the violations in the manual rosters.

As in the first case study, three sets of rosters were compiled: the “*Historical data*”, the “*Generated schedules with manual weights*”, and the “*Generated schedules with extracted weights*”. Again, the heuristic from Smet et al (2014) was used to automatically generate the latter two sets of rosters, following the procedure for manually constructing rosters, i.e. per quarter of a year. Each set consists of

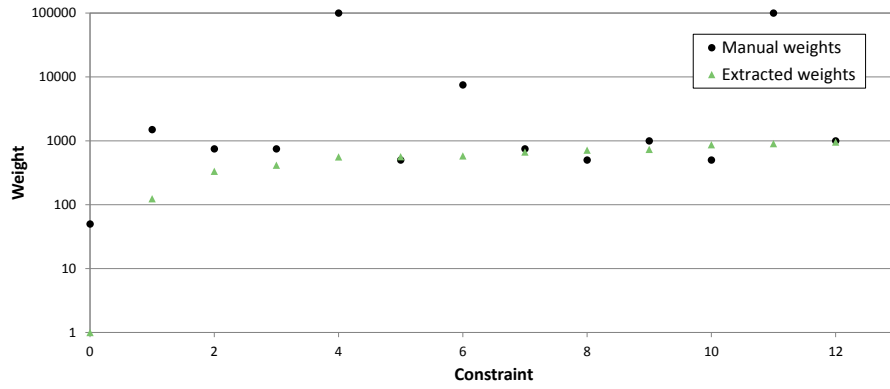
Table 2 Shifts in the geriatric medicine ward.

	Start	End	Duration	Break
Early shifts				
V01	6:45	15:15	8h30	30 min
V09	6:45	12:45	6h00	0 min
Day shifts				
D44	8:00	14:30	6h30	30 min
D77	8:00	14:30	6h30	22 min
V02	6:45	13:00	6h15	0 min
D10	7:00	15:06	8h06	30 min
D34	8:00	16:30	8h30	30 min
D70	7:00	15:30	8h30	30 min
D83	7:30	13:30	6h00	0 min
D175	7:30	15:36	8h06	30 min
D16	8:00	16:06	8h06	30 min
D82	7:15	15:45	8h30	30 min
Late Shifts				
A01	13:30	21:30	8h00	0 min
A11	15:30	21:30	6h00	0 min
A12	15:00	21:30	6h30	0 min
Night shift				
N03	21:15	7:00	9h45	0 min
Other shift				
LW8	7:30	11:30	4h00	0 min

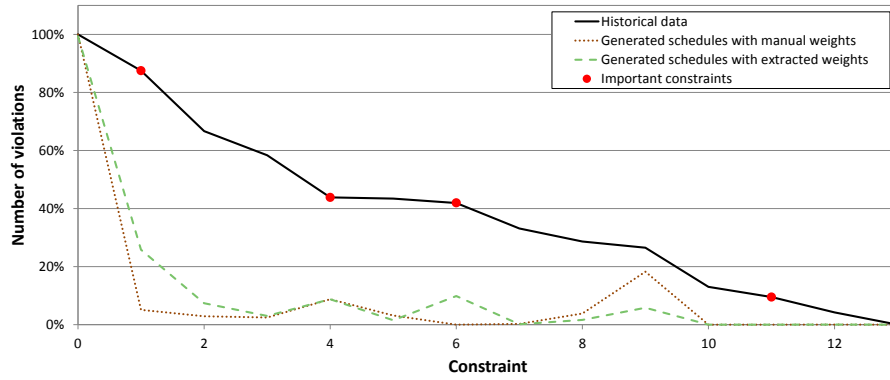
monthly rosters concerning Q1, Q2 and Q4. July, August and September were not considered, since they present special conditions due to reduced workload.

The evaluation by the head nurses partaking in the experiment was conducted in the following way. For each month in the considered periods, the rosters from the three sets were printed in a familiar format, bundled, and presented together with an evaluation form. The head nurses were asked to score, for their ward, each roster on a scale from 1 (very bad) to 10 (very good). To reduce bias, the different rosters were presented anonymously such that nurses were not aware whether they were evaluating a manual or automated roster and whether the latter was generated based on manual or extracted weights.

At the start of the evaluation, an explanation was given, briefly discussing the context of the experiments. The rosters were then (subjectively) evaluated and, relying on the experience of the head nurse, ranked based on their perceived quality. No time limit was imposed for the evaluation of the rosters. The experiment was



(a) Weights



(b) Violations

Fig. 4 Comparison of manually defined and extracted weights, and constraint violations in the different rosters for the GM ward.

conducted at a different time for each head nurse and there was no interaction between nurses during the experiment.

The outcome of the evaluation relies on the expertise of the head nurses to evaluate rosters for their ward. It was assumed that each head nurse is sufficiently experienced to correctly perform this task, even though the required effort might be greater for more complex wards such as OT. This assumption is reasonable, since constructing rosters is one of the main responsibilities of the head nurses, implying they are skilled at the task. Nevertheless, the results should be interpreted with care as it is possible that other head nurses would produce different evaluations for the same rosters.

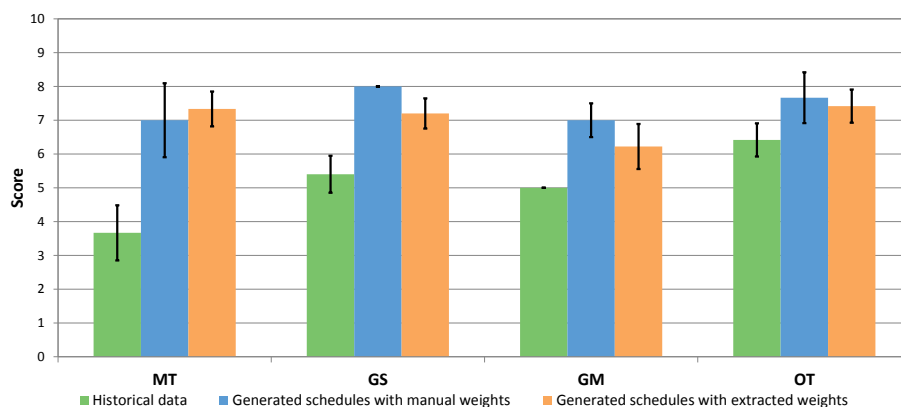


Fig. 5 Evaluation scores by human planners for different wards. The scores are on a scale of 0 to 10, higher denoting better.

Results

Prior to displaying the results of the evaluation forms, the data of the second case study are discussed.

Analysis

The values of the manual and extracted soft constraint weights are compared, as well as the number of constraint violations in the three sets of rosters. Results are only shown for the GM ward. The other wards presented comparable results.

The weights for the 14 constraints of the GM ward are shown in Figure 4a. Similar to the retrospective case study, manual weights are selected from a small set of discrete values, while the automatically extracted weights span a wider spectrum of values from 0 to 1000. Figure 4b shows the number of soft constraint violations in the three sets of rosters. The violations of the five constraints with the highest manually defined weights are indicated as “*Important constraints*”.

While the retrospective case study revealed a clear difference between the violations in the two sets of automatically generated rosters, the distinction was less obvious for the GM ward. Neither the manual, nor the extracted weights contribute to generating rosters that approximate the preferences of the head nurse, i.e. the violations in the historical data. Nevertheless, both sets of weights appear to generate strictly better solutions in terms of the number of soft constraint violations. One should therefore expect that the subjective evaluation by the nurses should reveal a comparable score for the two automatically generated rosters and that this score should exceed the one for the manual rosters.

The different outcomes of the two case studies can be partly explained by the recent transition of the GM ward. The manual weights had not significantly been re-adjusted by the head nurse from their initially defined values. Due to the low number of constraints in the second case study, the search algorithm succeeds in generating rosters that all Pareto-dominate the manual ones. Another possible explanation for the results is the low number of constraints, making it easier for the head nurse to prioritize and weigh constraints appropriately. Although it may be easier to express their relative importance in numerical form, it is certainly not trivial to actually take this importance into account when manually designing the rosters. Again, it is assumed that significant effort was spent in designing the manual rosters and that the head nurse was satisfied with their final quality.

Evaluation

The averages and standard deviations of the scores given by the head nurses are depicted in Figure 5. In all wards, the manually constructed rosters received the lowest scores of the three sets. This result can be read as a strong recommendation for automated rostering software, since 1) it appears to construct rosters that are more preferred than the manual ones and 2) it requires significantly less time to do so.

All head nurses made an important observation during the evaluation experiments. For constructing the manual rosters, it was occasionally necessary to consider a different set of constraints than the set defined in the rostering software, e.g. different coverage requirements or staff availability. The head nurses argued that there was no incentive to incorporate these short-term deviations in the software since they were often exceptional. All sets of rosters were evaluated based on the typical constraints defined in the software without considering the exceptional ones, thereby making the comparison with the manual rosters difficult (if not inapplicable here). This explains why for these wards, the AWE approach does not produce rosters that conform to the historical data. Moreover, it highlights an important drawback of automated rostering systems, namely their inability to deal with transient effects.

The two automatically generated sets of rosters show only a small difference in score. The observations confirm the prediction that the two sets of automated rosters evaluate comparably and have a better quality than the manual ones. In addition, the results do support the applicability of AWE: by automatically extracting weights from past data, the head nurse no longer needs to go through the time-consuming trial-and-error process of appropriately adjusting constraint weights.

As was mentioned in the experimental setup, these conclusions build upon the assumption that head nurses can correctly evaluate rosters for their ward. Even though the assumptions made were reasonable, the evaluations are still subjective, and could be different when other nurses were consulted.

Discussion

The results from both case studies deviate in their conclusions. In the retrospective case study AWE succeeded in producing rosters with a comparable distribution of constraint violations as in manually constructed rosters. This result is not observed in the prospective case study where neither of the two sets of automatically generated rosters follow the trend of the violation trade-off in the historical data. Based on knowledge obtained *after* discussions with the head nurses, the main cause is that the historical data from which AWE extracted weights considered a different set of constraints than the actual set used for generating the automated rosters. Nevertheless, through validation of the AWE approach with health care personnel, it was shown that the results are acceptable for practical execution. This conclusion was also formulated informally by the head nurses responsible for constructing the rosters, at the end of the experiment when the identity of the three sets of rosters was revealed. Both case studies illustrate how AWE succeeds in quantifying the head nurses' preferences regarding constraint priority. This strongly facilitates the initial investment phase, as it supports head nurses in defining the often subtle trade-off between constraint violations (Drake, 2014a). The evaluation scores are to be considered carefully as only four head nurses participated in the experiment, and more importantly, the evaluations are subjective.

Table 3 shows an overview of the benefits and drawbacks of current rostering practices in hospitals (fully manual rostering and automated rostering with manual transition), and the approach introduced in this article. It also highlights future trends, i.e. automated rostering with fully automated transition.

As pointed out, the AWE approach generates weights that conform to the importance of constraints in the historical data. Constraints in the historical data should therefore be identical to those used in the software for automated rostering. This assumption emphasizes the need for providing a relevant learning base for harnessing the full potential of automated weight extraction.

Finally, there is no reason to assume that the research results are limited to the wards considered in the case studies. The contributed AWE approach is general enough to be applied in yet other hospitals, countries, or even different types of organizations, provided sufficient relevant data are available. The approach simply reads through the relevant data, counting the occurrences and violations of a given

Rostering type and transition	Fully manual rostering	Automated rostering, manual transition	Automated rostering, semi-automated transition	Automated rostering, fully automated transition
Applied in	Current practice	Current practice	Proposed approach	Future trends
Benefits	+ considers implicit constraints + answers specific requests	+ saves time in rostering + higher departmental control	+ easier transition to automated + reduced time investment	+ minimal time investment + no human bias
Drawbacks	- labour intensive - requires experience - costs the department	- requires experience - unintuitive to translate experience - biased results	- requires data - difficulty handling transient effects - not suitable after drastic restructuring	- requires data - difficulty handling transient effects - may not capture complex constraints

Table 3 Comparison between different levels of automation in nurse rostering and the transition from manual to automated rostering

set of constraints and determines their weights, which are afterwards applied by the automated planner. All variables used in the weight extraction process (Equation 1) are formulated in an abstract manner, independent of the health care context.

Conclusions and outlook

Academic work on nurse rostering promises to automate the labor-intensive task of manually constructing rosters. However, in practice, it is hindered by the need to translate the practitioner’s experience in numerical form. The proposed AWE approach is capable of narrowing the research-application gap by automatically extracting soft constraint weights from past rosters. The methodology facilitates the manual configuration of automated rostering software for hospital wards, allowing for easier adoption of state of the art rostering approaches. Two case studies confirm the effectiveness of the new weight extraction approach, while also indicating the need for relevant historical data. Validation of the approach with health care practitioners responsible for constructing rosters indicated the difficulty of handling transient effects in AWE. Finally, initial feedback received from the head nurses who participated in the validation of the proposed approach, indicates that the potential time savings are substantial, as the iterative trial-and-error process to determine suitable weights is eliminated. AWE has been integrated in commercial software for automated rostering (Plan@SAGA, Tobania), and is deployed in various hospitals in Belgium.

Future research should focus on developing more intricate methods for automated weight extraction. For example, by taking into account the amount of violation, instead of only considering whether a constraint is violated or not. In addition, techniques can be explored for automated extraction of constraints from hospital regulations and departmental rules, thereby further bridging the gap between research and practice.

Acknowledgements The authors would like to thank SAGA Systems and the hospital staff for their cooperation. This research was carried out within the IWT 110257 project.

References

- Abdennadher S, Schlenker H (1999) INTERDIP – an interactive constraint based nurse scheduler. In: Proceedings of the First International Conference and Exhibition on The Practical Application of Constraint Technologies and Logic Programming
- Aickelin U, Dowsland K (2004) An indirect genetic algorithm for a nurse-scheduling problem. *Computers & Operations Research* 31(5):761–778
- Ásgeirsson E (2014) Bridging the gap between self schedules and feasible schedules in staff scheduling. *Annals of Operations Research* 218(1):51–69
- Ballestero E (2007) Compromise programming: A utility-based linear-quadratic composite metric from the trade-off between achievement and balanced (non-corner) solutions. *European journal of operational research* 182(3):1369–1382
- Berrada I, Ferland JA, Michelon P (1996) A multi-objective approach to nurse scheduling with both hard and soft constraints. *Socio-Economic Planning Sciences* 30(3):183–193
- Burke E, Curtois T, Post G, Qu R, Veltman B (2008) A hybrid heuristic ordering and variable neighbourhood search for the nurse rostering problem. *European Journal of Operational Research* 188(2):330–341
- Burke EK, De Causmaecker P, Vanden Berghe G, Van Landeghem H (2004) The State of the Art of Nurse Rostering. *Journal of Scheduling* 7(6):441–499
- Burke EK, Li J, Qu R (2012) Pareto-based optimization for multi-objective nurse scheduling. *Annals of Operations Research* 196(1):91–109
- Chankong V, Haimes Y (2008) Multiobjective decision making: theory and methodology. Dover Publications
- Charon I, Hudry O (2001) The noising methods: A generalization of some meta-heuristics. *European Journal of Operational Research* 135(1):86–101
- Coello C, Lamont G, Van Veldhuisen D (2007) Evolutionary Algorithms for Solving Multi-Objective Problems. Genetic and evolutionary computation, Springer Science & Business Media
- Drake RG (2014a) The nurse rostering problem: from operational research to organizational reality? *Journal of Advanced Nursing* 70(4):800–810
- Drake RG (2014b) The ‘robust’ roster: exploring the nurse rostering process. *Journal of Advanced Nursing* 70(9):2095–2106
- Eiben A, Van Hemert J (1999) SAW-ing EAs: Adapting the fitness function for solving constrained problems. In: Corne D, Dorigo M, Glover F, Dasgupta D,

-
- Moscato P, Poli R, Price KV (eds) *New Ideas in Optimization*, McGraw-Hill Ltd., UK, Maidenhead, UK, England, pp 389–402
- Ernst A, Jiang H, Krishnamoorthy M, Sier D (2004) Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research* 153(1):3–27
- Eveborn P, Rönnqvist M (2004) Scheduler—a system for staff planning. *Annals of Operations Research* 128(1-4):21–45
- Hulshof P, Kortbeek N, Boucherie R, Hans E, Bakker P (2012) Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems* 1(2):129–175
- Kelemci Ö, Uyar S (2007) Application of a genetic algorithm to a real world nurse rostering problem instance. In: *Proceedings of 9th International Conference on Enterprise Information Systems*, pp 474–477
- Kellogg DL, Walczak S (2007) Nurse scheduling: from academia to implementation or not? *Interfaces* 37(4):355–369
- Kudyba S, Perry T (2015) A data mining approach for estimating patient demand. *Health Systems* 4(1):5–11
- Marmion ME, Dhaenens C, Jourdan L, Liefoghe A, Verel S (2011) NILS: A neutrality-based iterated local search and its application to flowshop scheduling. In: Merz P, Hao JK (eds) *Evolutionary Computation in Combinatorial Optimization*, Lecture Notes in Computer Science, vol 6622, Springer Berlin Heidelberg, pp 191–202
- Mirrazavi S, Beringer H (2007) A web-based workforce management system for Sainsburys Supermarkets Ltd. *Annals of Operations Research* 155(1):437–457
- Parr D, Thompson J (2007) Solving the multi-objective nurse scheduling problem with a weighted cost function. *Annals of Operations Research* 155(1):279–288
- Petrovic S, Vanden Berghe G (2012) A comparison of two approaches to nurse rostering problems. *Annals of Operations Research* 194(1):365–384
- Silvestro R, Silvestro C (2000) An evaluation of nurse rostering practices in the national health service. *Journal of Advanced Nursing* 32(3):525–535
- Smet P, De Causmaecker P, Bilgin B, Vanden Berghe G (2013) Nurse rostering: A complex example of personnel scheduling with perspectives. In: Uyar AS, Ozcan E, Urquhart N (eds) *Automated Scheduling and Planning*, Studies in Computational Intelligence, vol 505, Springer Berlin Heidelberg, pp 129–153
- Smet P, Bilgin B, De Causmaecker P, Vanden Berghe G (2014) Modelling and evaluation issues in nurse rostering. *Annals of Operations Research* 218(1):303–326
- Tai B, Machin D (2013) *Regression Methods for Medical Research*. Wiley

Tourangeau AE, McGillis Hall L, Doran DM, Petch T (2006) Measurement of nurse job satisfaction using the McCloskey/Mueller Satisfaction Scale. *Nursing Research* 55(2):128–136